

УДК 004

ПРОБЛЕМЫ РАЗРАБОТКИ СИСТЕМЫ ДЛЯ ПОДДЕРЖКИ СОЦИАЛЬНЫХ ИССЛЕДОВАНИЙ В РФ

С.В. Романчуков

*Научный руководитель: О.Г. Берестнева, д.т.н., профессор
Томский политехнический университет, 634050, Россия, г. Томск, пр. Ленина, 30
E-mail: inoy@vtomske.ru*

This article is devoted to both algorithms and software for analytics subsystem of multifunctional web portal for social researches and notes some aspects of the analysis module design. Also it summarizes the elements of the Data Mining concept, used in the current system.

Keywords: data mining, social researches, analytics subsystem.

Ключевые слова: глубинный анализ данных, социальные исследования, аналитический модуль.

Введение

В современном обществе всё большую роль играет человеческий фактор, прямо зависящий от уровня и качества жизни населения, что порождает потребность в проведении разного рода социальных и психологических исследований. Получаемые в ходе выполнения данных работ объёмы данных требуют специальных программных средств для их обработки.

Формулирование требований к системе

Научно-учебной лабораторией информационных технологий в социальных и медицинских исследованиях разрабатывается тестовый портал «MultiTest». На настоящий момент работа идёт над аналитическим модулем, к которому выдвинуты следующие требования:

- 1) доступ через сеть Internet посредством браузера;
- 2) возможность обработки данных, хранящихся в файлах различных типов;
- 3) взаимодействие с наиболее распространёнными СУБД;
- 4) разделение прав доступа пользователей различных категорий;
- 5) размещение серверов на территории РФ в соответствии с законодательством [1].

Методы социологического исследования

Конкретный вид социологического исследования **обусловлен характером поставленной цели, выдвинутых задач**, в соответствии с чем различают **три основных вида** социологического исследования: разведывательное, описательное и аналитическое, и предполагает определенную последовательность этапов:

На первом этапе формируется программа исследования.

На втором этапе осуществляется сбор первичной информации.

На третьем производится цифровая обработка полученной информации.

На четвертом этапе проводятся анализ обработанной информации, формулирование выводов и рекомендаций. [2]

Очевидно, что применение сетевых информационных ресурсов ускоряет процесс проведения любого исследования.

Обработка результатов исследований

После представления результатов опросов и тестирований в той или иной шкале становится возможна их математическая обработка с использованием разнообразных статистических методов. Однако они имеют и свои ограничения, что касается, в первую очередь, методов параметрической статистики, в основе которых лежит предположение, что случайный вектор переменных образует некоторое распределение. Если это предположение не находит

подтверждения, следует воспользоваться непараметрическими методами математической статистики [3] в т. ч. DataMining – поиска в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний. [4]

Выделяют пять групп методов DataMining: ассоциация, последовательность, классификация, кластеризация и прогнозирование [5]. В случае социальных исследований на первый ряд выходят задачи кластеризации и классификации, т. к. с помощью *классификации* выявляются признаки, характеризующие группу, к которой принадлежит тот или иной объект, а с помощью кластеризации выделяют группы данных, которые заранее не были заданы, формируя новые продукты [6].

Для реализации данных методов широко используются программное обеспечение, которое можно разделить на три категории:

1. Универсальные пакеты (STATGRAPHICS, Statistica и т. д.).
2. Профессиональные пакеты (SPSS, SAS/IDS, BMDP).
3. Специализированные пакеты (EQS, ЭВРИСТА). [7].

Следует отметить, что широкие возможности в статобработке, предоставляемые данными инструментами, требуют от компьютера больших вычислительных ресурсов, кроме того часть из них поддерживает лишь отдельные операционные системы или имеет существенные проблемы при работе с сетью, поддерживая только работу в локальной сети или сталкиваясь с большим количеством ограничений по размеру и структуре обрабатываемых данных, величине обмена данных с сервером и т. д. [8]. Исследования, связанные с обработкой персональных данных, сталкиваются также с законодательными ограничениями, что исключает использование зарубежных серверов.

Главным же препятствием на пути освоения этих программ является время, которое необходимо затратить на обучение. Именно из-за недостатка знаний у пользователя мощь статистических пакетов зачастую не используется даже в половину [9].

Немаловажным недостатком является и высокая цена таких программных продуктов.

Всё это обращает нас к необходимости разработки программных решений, ориентированных на нужды реально существующих исследовательских групп, совместно с их участниками.

Список литературы

1. Ст. 18, ст. 22 Федерального Закона РФ «О персональных данных» от 27 июля 2006 г. N 152-ФЗ после изменений в соответствии с ФЗ N 142-ФЗ от 4 июня 2014 г.).
2. Виртуальный учебно-методический комплекс «Политическая наука» [Электронный ресурс] URL: http://read.virmk.ru/s/SANZ_SOC/g-014.htm доступ свободный.
3. Лаборатория статистических исследований Кубанского государственного университета, Обзор методов статистического анализа данных [Электронный ресурс] URL: <http://www.statlab.kubsu.ru/node/4>.
4. Григорий Пятецкий-Шапиро, DataMining и перегрузка информацией // Вступительная статья к книге: Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. 3-е изд. перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с. С. 13.
5. Типы закономерностей, выявляемых методами DataMining [Электронный ресурс]. – URL: http://fsecrets.ru/2010/10/типы_закономерностей_выявляемых_мет/ доступ свободный.
6. Data Mining: Concepts, Models, Methods and Algorithms / MehmedKantardzic – New Jersey, 2011. – С. 249–253.
7. Компьютерные методы обработки данных: обзор статистических пакетов [Электронный ресурс] URL: <http://denisvolkov.com/wp-content/uploads/2011/03/KMOD-0.pdf> доступ свободный.
8. Официальный сайт разработчика Statgraphics. Условия предоставления доступа к среде Statgraphics Online [Электронный ресурс] URL: http://statgraphics.com/statgraphics_online.htm доступ свободный.
9. Обзор статистических программ [Электронный ресурс] URL: <http://www.sciencefiles.ru/section/46/> доступ свободный.